GENE CODING FOR AN ACYLTRANSFERASE OF OIL SEED RAPE, AND USES THEREOF

The invention relates to the identification and to the cloning of a gene encoding an acyltransferase, and to uses thereof.

In plants, glycerolipids (glycolipids, phospholipids and triacylglycerides) constitute the major portion of lipids. Their common precursor is *sn*-1,2-diacylglycerol-3-phosphate, or phosphatidic acid (PA), resulting from the esterification of the *sn*-1 and *sn*-2 positions of glycerol-3-phosphate (G3P) with fatty acids. *SN*-glycerol-3-phosphate acyltransferase (GPAT) (E.C. 2.3.1.15) catalyzes the acylation of the *sn*-1 position of G3P so as to form *sn*-1-acylglycerol-3-phosphate, or lysophosphatidic acid (LPA). The LPA is then used as a substrate by 1-acyl-*sn*-glycerol-3-phosphate acyltransferase, or lysophosphatidic acyltransferase (LPAAT) (E.C.2.3.1.51), which acylates the *sn*-2 position of glycerol. In the synthesis of triacylglycerides, which constitute most of the storage lipids, a third enzyme, *sn*-1,2-diacylglycerol acyltransferase, or diacylglycerol acyltransferase (DAGAT), is involved, which catalyzes the acylation of the *sn*-3 position.

Plant lipids are currently used in very varied fields, from foods to the chemical industry, and it is desirable to have plants which produce lipids specifically suited to the use envisaged. With this aim, attempts are being made in particular to modify the fatty acid composition of the glycerolipids, and in particular of the triacylglycerides.

For example, in the case of rapeseed (*Brassica napus*), plants having the lowest possible erucic acid content are used for edible oil. On the other hand, a high erucic acid content is sought in plants which produce oils intended for industrial use.

The fatty acid composition of the glycerolipids depends essentially, firstly, on the quantitative and qualitative distribution of the fatty acids produced by the plant and, secondly, on the substrate specificity of the acyltransferases with respect to these fatty acids. In order to control this composition, it has been proposed to act separately or jointly on these two factors, by intervening:

- at the level of biosynthesis of the fatty acids, in order to promote, or on the contrary inhibit, the production of one or more specific fatty acids and, optionally, to induce the synthesis of new fatty acids;
- at the level of acylation of the G3P, in order to modify its specificity in the desired way.

In the case of rapeseed, the most erucic acid-rich varieties currently available produce an oil in which erucic acid represents at most 50 to 60% of the total fatty acids. Analysis of the triacylglycerides of the seeds derived from these varieties has shown that this acid is present virtually exclusively at the *sn*-1 and *sn*-3 positions; this selective distribution has been attributed to the substrate specificity of rapeseed LPAAT, which excludes very-long-chain (>C20) fatty acids; this limits the erucic acid content of the triacylglycerides of seeds from rapeseedseed to a theoretical maximum threshold of 66% of the total fatty acids.

With the aim of avoiding this limitation, the gene of a *Limnanthes alba* LPAAT capable of incorporating erucic acid at the *sn*-2 position has been expressed in the seeds of a variety of rapeseed having a high erucic acid content. However, although incorporation of erucic acid at the *sn*-2 position was effectively observed in the triacylglycerides of the seeds of this transgenic rapeseed, this incorporation remained low; in addition, the total amount of erucic acid incorporated into these triacylglycerides was no greater than that of the nontransformed control plants [LASSNER et al., Plant Physiol. 109:1389-1394, (1995)].

This result may be due, besides the possibility of a limiting production of erucic acid, to poor specificity of the exogenous *Limnanthes* LPPAT which might, besides erucic acid, also incorporate oleic acid, and also to the existence of competition between the exogenous LPAAT activity and endogenous LPAAT activity of the rapeseed, which it would be necessary to inhibit in order to increase the incorporation of erucic acid.

To date, only a small amount of information has been available concerning the enzyme(s) responsible for LPAAT activity in rapeseed. LPAATs are in fact membrane-bound enzymes which are difficult to purify in active form.

With the exception of coconut LPAAT [KNUTZON et al., Plant Physiol. 109:999-1006, (1995)], which has been purified from the membranes of the albumen, and the gene of which has subsequently been isolated by screening a cDNA library, the plant LPAATs already identified have mostly been characterized using techniques of molecular genetics. They are maize LPAAT [BROWN et al., Plant Mol. Biol., 26:211-223, (1994)], and Limnanthes LPAATs [BROWN et al., Plant Mol. Biol., 29:267-278, (1995); HANKE et al., Eur. J. Biochem. 232:806-810, (1995)].

In order to enable the control of the acylation at the *sn*-2 position to be improved, the inventors have undertaken to characterize the enzyme(s) involved in LPAAT activity in rapeseed.

They have thus succeeded in isolating a DNA sequence from *Brassica napus* encoding a functional plastidial LPAAT; this LPAAT will hereinafter be named BAT2 (Brassica AcylTransferase 2).

A nucleic acid sequence comprising the sequence encoding BAT2 is represented in the attached sequence listing under the number SEQ ID NO: 1, and the deduced polypeptide sequence is represented under the number SEQ ID NO: 2.

2 ATG codons which may constitute potential translation initiation sites are present on the sequence SEQ ID NO: 1; the polypeptide of 344 amino acids starting at the methionine residue at position 16 of the sequence SEQ ID NO: 2 is sufficient for the LPAAT activity.

The analysis of the amino acid sequence of BAT2 using the pSORT program suggests the presence of a signal sequence included in the 95 N-terminal amino acids of the sequence SEQ ID NO: 2. This signal sequence is involved in directing the BAT2 LPPAT into the plastidial membrane.

The sequence of the active mature protein is included in the 279 C-terminal amino acids.

The comparison, using the BLASTX2 program [GISH et al., Nat. Genet., 31:266-272, (1994)], between the BAT2 peptide sequence and the LPAAT peptide sequences previously known reveals very weak homology (a maximum of 20% identity) when the comparison is carried out over the entire sequence.

Over certain regions of the sequence, greater homology is observed. Figure 1 represents the alignment of the sequence 187-302 of BAT2 with the sequences of the LPAATs having the strongest homology. The most significant scores are observed with:

- the product of the *S. cerevisiae SLC1* gene (P33333) [NAGIAC et al. J. Biol. Chem., 268:22145-22163, (1993)] : 32% identity and 51% equivalence, over an alignment of 204 amino acids;

- the microsomal LPAAT of the seeds from Limnanthes (Q42870) [HANKE et al., Eur. J. Biochem., 232:806-810, (1995); LASSNER et al., Plant Physiol. 109:1389-1394, (1995); BROWN et al., Plant Mol. Biol., 29:267-278, (1995)] : 30% identity and 54% equivalence, over an alignment of 182 amino acids;

- the coconut endosperm LPAAT (Q42670) [KNUTZON et al., Plant Physiol. 109:999-1006, (1995)] : 31% identity and 47% equivalence, over an alignment of 229 amino acids;

- the hypothetical *Synechocystis* LPAAT (P74498) : 30% identity and 55% equivalence, over an alignment of 143 amino acids;

- the *E. coli* plsC protein (P26647) : 31% identity and 50% equivalence, over an alignment of 115 amino acids.

A subject of the present invention is a nucleic acid fragment comprising:

a) a sequence encoding a plant LPAAT, the peptide sequence of which shows at least 20%, preferably at least 30%, and advantageously at least 50 to 95%, identity with the sequence SEQ ID NO: 2; and/or

b) a sequence complementary to the coding sequence a) above.

According to a preferred embodiment of the present invention, said coding sequence encodes the polypeptide of sequence SEQ ID NO: 2.

The invention also encompasses fragments of more than 20 bp, and preferably of more than 30 bp, which are fragments of a coding sequence as defined above, or which are capable of hybridizing specifically, under stringent conditions, with said sequence. This includes in particular the fragments of any sequence encoding the polypeptide SEQ ID NO: 2, or of the sequence complementary thereto, with the exception of the fragments consisting of an oligonucleotide encoding one of the following peptide sequences (1-letter code):

FPEGTRS;

PFKKGA;

which are common to LPAATs having previously known sequences, or of the fragments complementary to said oligonucleotide.

Nucleic acid fragments in accordance with the invention can in particular be used as primers and/or probes, to detect and clone sequences encoding plastidial LPAATs, from rapeseed or from other plants, and also sequences encoding LPAATs from rapeseed or from other species, in particular from crucifers, which

are expressed in cellular compartments other than plasts, in particular the endoplasmic reticulum.

The analyses carried out by Southern transfer and by RFLP, using the *BAT2* cDNA, labeled with ^{32}P, as a hybridization probe, under stringent conditions, show the presence, in the rapeseed genome and also in the *A. thaliana* genome, of at least 2 homologous copies of the *BAT2* gene, and imply that this gene is part of a multigene family comprising 4 members.

A subject of the present invention is also:

- the recombinant vectors resulting from the insertion of at least one nucleic acid fragment in accordance with the invention into a suitable vector; advantageously, they are expression vectors in which the nucleic acid fragment in accordance with the invention is inserted under the transcriptional control of regulatory sequences (such as a promoter and/or terminator) which are functional in a host cell in which the expression of said fragment is desired.

- the host cells, which may be prokaryotic or eukaryotic, and the multicellular organisms, in particular plant cells and plants, transformed with at least one nucleic acid fragment in accordance with the invention.

The invention also encompasses the recombinant LPAAT, or the fragments of recombinant LPAAT, resulting from the expression, in a host cell, of the sequence encoding said LPAAT or said fragment, which sequence is carried by a nucleic acid fragment in accordance with the invention. The recombinant LPAAT in accordance with the invention, or the fragments thereof, can, for example, be used to produce anti-LPAAT antibodies which enable cDNA expression libraries to be screened in the context of detecting and cloning other LPAATs.

Nucleic acid fragments in accordance with the invention can advantageously be used, in the sense or antisense orientation, to produce transgenic plants, in particular from rapeseed or from other oil-producing plants, in order to regulate the LPAAT activity in the

plant thus transformed, and to act on the fatty acid composition of the lipids, and in particular of the triacylglycerides, produced by this plant.

The present invention also encompasses the transgenic plants produced in this way.

These plants can be produced using the conventional techniques, which are known in themselves, of plant transgenesis. Depending on the use envisaged, a nucleic acid sequence in accordance with the invention can be placed under the control of an inducible promoter or of a constitutive promoter, of a ubiquitous promoter or of a tissue-specific promoter. These plants can also contain other transgenes, preferably transgenes derived from genes involved in lipid biosynthesis.

It is possible in particular to produce:

- transgenic plants expressing at least one sequence in accordance with the invention encoding a functional LPAAT, instead of and in place of one or more sequences encoding endogenous LPAATs, or in addition to these sequences;

- transgenic plants expressing at least one sequence in accordance with the invention in antisense orientation, in order to inhibit the expression of the homologous endogenous LPAATs and thus promote the activity of other LPAATs, of endogenous or exogenous origin.

For example:

- in order to produce transgenic rapeseed plants having a high erucic acid content, the rapeseed plant can be cotransformed with, firstly, a DNA sequence encoding an LPAAT which preferentially incorporates the erucic acid at the *sn*-2 position, such as the *Limnanthes alba* LPAAT [LASSNER et al., (1995), abovementioned publication], and, secondly, a nucleic acid sequence in accordance with the invention in antisense orientation, in order to inhibit, at least partially, the production of endogenous LPAAT which

competes with the activity of the exogenous *Limnanthes* LPAAT;

- in order to increase the overall triglyceride content of the seeds, the rapeseed plant can be transformed with a DNA sequence in accordance with the invention which encodes a plastidial LPAAT deleted of its sequence for being directed into plasts;

- in order to increase the saturated fatty acid content, in particular the palmitic acid content, the rapeseed plant can be cotransformed with a DNA sequence in accordance with the invention which encodes a plastidial LPAAT deleted of its sequence for being directed into chloroplasts, and with one or more genes for ACP-thioesterases which preferentially use palmitoyl-ACPs as substrates.

The present invention will be more clearly understood with the aid of the further description which follows, which refers to nonlimiting examples illustrating the identification and cloning of a gene encoding the *Brassica napus* BAT2 LPAAT.

EXAMPLE 1: ISOLATION AND CHARACTERIZATION OF THE cDNA OF A RAPESEED LPAAT

In order to search for the presence of genes encoding LPAATs, a rapeseed immature embryo cDNA library was screened by heterologous complementation of the mutation of the *plsC* gene of the *E. coli* strain JC201 [COLEMAN, J. Biol. Chem., 265:17215-17221, (1990)]. This point mutation confers a heat-sensitive phenotype on the JC201 mutants, due to the inactivation at high temperature of the LPAAT encoded by the *plsC* gene. These mutants grow well at 30°C, with difficulty at 37°C, and not at all at 42-44°C.

For the screening, a phagemid library derived from 2×10^6 clones taken from a rapeseed immature embryo cDNA library and cloned into the vector lambda ZAPII (STRATAGENE), was constructed using the "ExAssist" kit (STRATAGENE).

The bacteria are transformed with these phagemids by electroporation, and then cultured on LB

agar, in the presence of ampicillin and IPTG (isopropyl-β-D-galactothiopyranoside). The bacteria which grow at 42°C are selected. The plasmid DNA of the clones capable of growing at 42°C was analyzed by PCR in order to determine the size of the insert. After 3 cycles of transformation followed by selection, approximately 85% of the clones contain an insert of approximately 1.2 kb. The sequencing of the ends of the inserts of 4 of these clones shows that they are identical. One of these clones, named pBAT2, was entirely sequenced.

EXAMPLE 2: NUCLEOTIDE SEQUENCE OF *BAT2* AND DEDUCED PEPTIDE SEQUENCE.

The cDNA of the pBAT2 clone comprises a 1155 bp sequence followed by an 18 residue poly(A) tail. This sequence comprises a single open reading frame, corresponding to a 351 amino acid polypeptide, which represents a fusion protein, comprising 344 amino acids of the pBAT2 sequence, and a portion of the β-galactosidase sequence from the cloning vector. The sequence which is represented on the attached sequence listing under the number SEQ ID NO: 1 also comprises a portion of the genomic sequence (nucleotides 1 to 79 of the sequence SEQ ID NO: 1) located upstream of the pBAT2 cDNA sequence.

2 ATG codons which may constitute potential translation initiation sites were located on the sequence SEQ ID NO: 1; if the first one of them (position 58 of the sequence SEQ ID NO: 1) is used, the product of translation of the sequence SEQ ID NO: 1 is a 359 amino acid polypeptide, the theoretical molecular weight and pI of which are, respectively, 39.6 kDa and approximately 9.8; this polypeptide is represented on the attached sequence listing under the number SEQ ID NO: 2. If the translation initiation takes place at the 2nd ATG codon (position 103 of the sequence SEQ ID NO: 1), the translation product is a 344 amino acid polypeptide, the theoretical molecular weight of which is approximately 37.9 kDa.

The analysis of the BAT2 amino acid sequence using the pSORT program suggests the presence of a signal peptide of approximately 80 to 95 residues. This potential signal sequence is rich in serine, in alanine, in valine and in basic amino acids, which is characteristic of sequences for directing toward the chloroplast membrane.

The analysis of the polypeptide sequence also indicates the presence of two potential transmembrane domains, located, respectively, between amino acids 124 to 140, and 219 to 235.

The consensus sequences of the LPAATs (FPEGTRS and PFKKGA) are located, respectively, at positions 273-279 and 286-291 of the sequence SEQ ID NO: 2; a sequence corresponding to the consensus sequence NHXXXXD, which is conserved in all the membrane-bound acyltransferases known to date, is located at positions 202-208 of the sequence SEQ ID NO: 2.

EXAMPLE 3: ENZYMATIC ACTIVITY OF THE PROTEIN ENCODED BY THE pBAT2 cDNA INSERT

In order to verify that the protein encoded by the pBAT2 insert effectively had an LPAAT activity, the capacity of this protein to incorporate oleic acid or palmitic acid at the *sn*-2 position of LPA was tested.

The *E. coli* strain JC201 transformed with pBAT2 and, by way of controls, the *E. coli* strain JC201 either nontransformed or transformed with the vector (pBSK) lacking the *BAT2* cDNA insert are cultured at 30°C until an optical density of 0.5 is reached.

After induction with IPTG and culturing for 3 h at 30°C, the bacteria are lyzed and fractionated and the LPAAT specific activity is measured, on the crude membrane extracts, in the presence of oleoyl-CoA(1-^{14}C) or of palmitoyl-CoA(1-^{14}C), according to the protocol of CAO et al. [Plant Physiol., 9:1199-1206, (1990)].

The results are shown in Table I below, which gives the specific activity, in pmol of phosphatidic acid formed/mg of protein/hour.

TABLE I

CULTURE	SUBSTRATE	
	Oleoyl-CoA(1-^{14}C)	Palmitoyl-CoA(1-^{14}C)
JC201	1.86	3.74
JC201+pBSK	1.65	1.59
JC201+pBAT2	5.8	11.06

The membrane extracts of the culture transformed with pBAT2 exhibit an LPAAT activity which is greater than that of the membrane extracts obtained from the nontransformed culture or the culture transformed with the vector pBSK, which shows that the LPAAT activity is effectively restored by the product of translation of the pBAT2 insert.

EXAMPLE 4: CELLULAR LOCATION OF BAT2

The plastid location suggested by the BAT2 sequence analysis was verified by testing the capacity of isolated pea chloroplasts to import BAT2.

With this aim, the cDNA of the pBAT2 clone was transcribed *in vitro*, using T3 RNA polymerase; the transcript is translated in the wheat germ acellular system, in the presence of ^{35}S-labeled methionine. A translation product of approximately 40 kDa is thus obtained. This product is incubated with isolated pea chloroplasts. After incubation, the chloroplasts are treated with protease and fractionated, according to the protocol described by BROCK et al. [Plant Mol. Biol. 23(4), 717, (1993)], and the various fractions are analyzed by electrophoresis in order to search for the ^{35}S-labeled product.

The results of this analysis show that the *BAT2* translation product is imported into the pea chloroplasts and cleaved into a 32 kDa protein which is essentially located in the membrane fraction, and an 8 kDa signal peptide.

These results confirm that the BAT2 protein is indeed synthesized with a signal peptide, the role of which is to import the protein into the membrane of the chloroplasts. The precursor has an apparent mass of

approximately 40 kDa and the signal peptide has an apparent mass of approximately 8 kDa.

EXAMPLE 5 : LOCATION OF THE EXPRESSION OF THE BAT2 GENE

The expression of the *BAT2* gene was studied in various organs of *B. napus* and of *A. thaliana*.

The study was carried out by Northern transfer, using a probe corresponding to the *BAT2* coding sequence, on total RNAs of stems, roots, leaves, flowers, seeds undergoing development [28 DAP (days after pollination)] and dry seeds.

In each of the *B. napus* and *A. thaliana* tissues tested, hybridization of the probe with a transcript of approximately 1.3 kb in *B. napus*, and with a transcript of approximately 1 kb in *A. thaliana*, is observed. The strength of the hybridization signal is similar in all the tissues, including the nonphotosynthetic tissues containing plasts other than chloroplasts. The observation of hybridization in mature seeds indicates, in addition, that the BAT2 message remains stable during maturation of the seed.